DOI: https://doi.org/10.15276/hait.09.2026.21
UDC 004.8

Enhancing the accuracy and interpretability of real estate price predictions using machine learning methods

Dmitriy E. Kravtsov[1)]
ORCID: https://orcid.org/0009-0005-0305-544X; dmkravtsov@gmail.com. Scopus Author ID: 59197585500
Nikolay I. Poletaev[1)]
ORCID: https://orcid.org/0000-0002-1340-582X; Poletaev@ukr.net. Scopus Author ID: 6603897743
Piotr Gorzelańczyk[2)]
ORCID: https://orcid.org/0000-0001-9662-400X; Piotr.gorzelanczyk@ans.pila.pl. Scopus Author ID: 22979733400
Edgar Sokolovskij[3)]
ORCID: https://orcid.org/0000-0002-0770-4225; edgar.sokolovskij@vilniustech.lt. Scopus Author ID: 55903092900
[1)] Odesa National Maritime University, 34, Mechnikov Str. Odesa, 65029, Ukraine
[2)] Stanislaw Staszic State University of Applied Sciences in Piła, 10, Podchorążych Str. Piła, 64-920, Poland
[3)] Vilnius Gediminas Technical University, 11, Saulėtekio al. Vilnius, LT-10223, Lithuania

ABSTRACT

Relevance: stems from the increasing need for machine learning models in real estate that are not only accurate but also interpretable, since practical applications require a clear understanding of the factors influencing predictions, particularly in decision-making related to urban development and property valuation. **Aim:** the aim of the article is to develop and evaluate an approach that simultaneously improves predictive performance and enhances the interpretability of machine learning models when working with heterogeneous data sources. **Objectives:** the objectives include analysing multiple feature groups, namely textual property descriptions, spatial indicators such as distances to key infrastructure objects, and visual characteristics derived from satellite-based night-time illumination data, as well as assessing their combined impact on model performance. **Methods:** the study is based on gradient boosting over decision trees using the Light Gradient Boosting Machine algorithm, the construction of spatial features through distance-based metrics, text processing techniques, and interpretation tools based on Shapley values and partial dependence analysis to reveal feature influence. **Scientific novelty:** the novelty lies in integrating heterogeneous features of different origin within a unified modelling framework and combining complementary interpretation techniques to identify nonlinear relationships and interaction effects. **Practical significance:** the proposed approach can be applied in automated valuation systems, urban analytics, and decision-support tools, providing a more transparent understanding of price formation mechanisms. **Results:** the results show that the proposed modelling approach based on the Light Gradient Boosting Machine provides high predictive performance. Starting from the baseline feature set, the model achieves an RMSE of 374 and an R^2 of 0.818. The integration of heterogeneous feature groups further improves the model performance, reducing the RMSE to 348 and increasing the R^2 to 0.839. It is also evident that textual and visual features play a noticeable role, as they help capture nonlinear patterns and threshold-like effects that are difficult to detect otherwise. **Conclusions:** overall, the proposed approach not only increases predictive accuracy but also makes the model behavior easier to interpret, leading to a more transparent and reliable analysis of real estate prices and supporting its use in practical applications.

Keywords: Machine learning; deep learning; mathematical model; statistical research; probability; correlation; forecast; neural networks; data mining; natural language processing; transformers; text embeddings; interpretability; transport; explainable artificial intelligence; model interpretability; interpretable machine learning; feature engineering; gradient boosting; decision support systems

***For citation*: Kravtsov D. E., Poletaev N. A., Gorzelańczyk P., Sokolovskij E., "Enhancing the accuracy and interpretability of real estate price predictions using machine learning methods". *Herald of Advanced Information Technology.* 2026; Vol.9 No.3: 321–335. DOI: https://doi.org/10.15276/hait.09.2026.21**

INTRODUCTION

The interpretability of machine learning models is becoming one of the key requirements in the tasks of mass real estate appraisal [1], [2]. For businesses, investors, developers, and governmental bodies, it is important not only to obtain an accurate price prediction but also to understand which factors drive it. Machine learning approaches have also been successfully applied to complex analytical problems such as anomaly and intrusion detection in computer networks, where models must identify hidden patterns in high-dimensional data spaces [3]. Despite the high accuracy of modern algorithms such as XGBoost, LightGBM, and neural networks, their internal logic often remains opaque, which reduces trust in the results and complicates their practical application [4]. This is especially critical when the model includes features of diverse nature – from objective property characteristics to subjective descriptions and indirect indicators. Examples of such features include the number of public transport stops, the density of fast-food restaurants within a given radius, the presence of the words "gym" or "swimming pool" in the property description, and the distance to centres of nocturnal illumination identified from satellite imagery [5], [6], [7]. Without transparent interpretation of the influence of these factors, it is difficult to adapt the model to market changes, identify sources of error, and employ it in sensitive domains such as mortgage lending, pricing, or urban planning. Machine learning approaches are also widely applied in decision-support systems for economic and investment analysis, including risk assessment models that assist investors in evaluating complex project parameters and forecasting potential outcomes [8]. Machine learning and soft-computing techniques are also used to analyse complex software metrics and predict system properties such as software reusability, demonstrating the ability of data-driven models to capture relationships between multiple heterogeneous parameters [9]. Machine learning techniques are widely applied for analysing complex technical, socio-economic, and biomedical systems and forecasting dynamic processes. In particular, they are used for cyber-attack detection and network security tasks, including DDoS analysis and mitigation [10], [11], as well as for real-time prediction of operational parameters of industrial equipment and modelling of its behaviour based on sensor data [12]. They are also applied to the prognosis of ageing phenomena in critical infrastructure systems [13] and to support data-driven decision-making in digital markets through intelligent information systems [14]. In the bio-medical domain, machine learning enables early prediction and analysis of chronic diseases based on medical datasets [15].

Recent studies demonstrate the growing role of machine learning in analysing complex heterogeneous data across different domains. In particular, semantic analysis of scientific publications enables automated knowledge extraction and identification of thematic structures in large document collections [16]. Advanced data-processing frameworks are also used for analysing complex trajectory datasets and detecting hidden dynamic patterns in navigation data [17]. At the same time, natural language processing methods and context-aware text embeddings are widely applied for modelling semantic relationships in textual data and improving information retrieval tasks [18], [19]. In addition, systematic approaches to data preprocessing and feature selection have been proposed to improve machine learning model performance and reduce computational costs [20].

In the context of real estate valuation, machine learning methods are increasingly applied to predict property prices and to analyse the contribution of various factors influencing rental value. The integration of heterogeneous data sources, including textual descriptions, spatial characteristics, and environmental indicators, makes it possible to capture complex relationships that cannot be identified using traditional approaches. At the same time, the use of interpretable machine learning techniques enables not only accurate prediction but also a detailed understanding of the underlying drivers of price formation.

The inclusion of a broad range of studies covering various aspects of data analysis, text processing, and machine learning model development is motivated by the need to demonstrate the interdisciplinary nature of the problem under consideration and to justify the relevance of the integrated approach adopted in this study.

The aim of this work is to develop and test an integrated approach to improving both the accuracy and interpretation of the impact of heterogeneous features — numerical, textual, geospatial, and visual — on predicted rental property values.

To achieve this aim, the following tasks were addressed:

1) the construction and enrichment of the initial rental property dataset for the city of Houston with additional features (textual, geospatial, and visual) to improve the performance of the machine learning model;

2) training and hyperparameter optimisation of the LightGBM ML model using recursive feature elimination to obtain the optimal feature set;

3) interpretation of the contribution of heterogeneous features to the machine learning model results, with emphasis on the use of SHAP and PDP;

4) evaluation of the universality of the obtained patterns and their transferability to other markets.

RELATED WORKS

In recent years, methods of interpreting machine learning models have become an important research direction within Explainable AI. General reviews systematise post-hoc interpretation approaches and highlight their applicability for analysing black-box models [4]. In the context of real estate valuation, interpretable machine learning is increasingly used to explain price determinants

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

and improve trust in automated valuation models [1], [21], [22].

Among the most widely used interpretation techniques are SHAP, PDP, ALE, and LIME. SHAP is commonly applied to estimate the contribution of individual features and to analyse their global importance [23]. Partial Dependence Plots are used to investigate the effect of feature variation on model predictions and to identify nonlinear relationships [24]. Additional approaches combine Shapley values with partial dependence functions to improve the robustness of interpretation results.

Recent studies show that machine learning models can improve valuation accuracy while providing insights into price formation mechanisms. Ensemble-based and explainable approaches have been successfully applied to property valuation in urban markets [25]. At the same time, studies of residential rental markets demonstrate how interpretable models can reveal internal pricing structures and feature interactions [26]. However, most existing works focus on a limited set of interpretation tools or rely on a single dominant data type, which restricts the analysis of relationships between heterogeneous features.

Another important limitation concerns the transferability of identified dependencies. The influence of real estate factors may vary significantly depending on the city and socio-economic conditions. For example, proximity to transport infrastructure may increase property value in one region but reduce it in another due to noise or congestion effects [7], [26]. This raises the question of how stable such dependencies are across different markets [27], [28].

Thus, the current literature confirms the importance of interpretable machine learning for real estate valuation but also highlights the need for approaches that integrate heterogeneous data sources and complementary interpretation techniques. The present study addresses this gap by combining numerical, textual, geospatial, and visual features within a unified modelling framework and by jointly applying SHAP and PDP to analyse nonlinear effects and feature interactions.

MACHINE LEARNING TOOLS FOR DATA INTERPRETATION

Initial data

As part of the study, a detailed analysis and preprocessing of the initial dataset were carried out, comprising 9,260 real estate properties and 10 features, one of which is the target variable “Price” (Table 1). The data relate to properties located in the city of Houston (USA) and were obtained from publicly available listings on the Redfin platform [29], which is widely used in real estate analytics. The dataset was formed by selecting residential rental listings with complete information on key attributes, including price, location, and property characteristics. To ensure data reliability, a series of preprocessing steps was performed, including the removal of duplicate entries, handling of missing values, and elimination of outliers and anomalous observations.

The dataset was collected during July 2023 and reflects the spatial, economic, and temporal characteristics of the residential rental market in Houston, Texas, USA. Therefore, the quantitative relationships identified in this study may vary when transferred to other cities, countries, or time periods, which should be taken into account when generalising the proposed approach.

Prior to analysis, each feature was checked for completeness and consistency, which allowed for maintaining the quality, reliability, and interpretability of the data used in the study.

Table 1. **Description of Basic Features**

Feature	Description
Latitude	Geographic latitude
Longitude	Geographic longitude
YearBuilt	Construction year
Beds	Number of bedrooms
Baths	Number of bathrooms
BuildingSize	Building area (sq ft)
LotSize	Land area (sq ft)
PostalCode	Postal code
Description	Property text description
Price	Rental price (USD)

***Source:* compiled by the authors**

Methods of Feature Engineering

Features Derived from the Textual Description of Real Estate Objects

Below is an example of a real estate property description presented in textual format. In addition to basic characteristics, it may contain details of layout, finishing, and the infrastructure of the complex, such as:

“*These apartments feature an open concept living/dining room and a kitchen with granite countertops, subway tile backsplash, and stainless-steel appliances with a dishwasher. Features include central A/C, central heat, and ceiling fans. The bedroom(s) have carpet flooring, and the rest of the home has hardwood floors. Select units offer nine-foot ceilings, built-in desks, shelving, and a private* *balcony. The community provides elevators, wheelchair access, a saltwater rooftop pool with sundeck, rooftop tennis court, and more."*

To transform textual descriptions into informative features suitable for the model, the following step-by-step process was implemented:

• tokenisation: the text was split into individual words (tokens) to enable analysis at the lexeme level;

• removal of stop words: frequently used but low-value words (e.g., conjunctions, prepositions) were excluded to reduce noise;

• lemmatization: words were converted to their base form, allowing different morphological variants of the same term to be unified;

• vectorisation using the TF-IDF method [30]: texts were converted into numerical format, where each word was assigned, a weight depending on its significance for a specific property and its rarity in the rest of the dataset;

• feature construction: TF-IDF weights of key words (e.g., pool, laundry, garage) were added to the dataset as additional features, reflecting the marketing and subjective characteristics of properties.

This approach enabled the model to incorporate textual descriptions as a full-fledged source of information influencing the final valuation of real estate [31]. In particular, based on text analysis, indicator features were created for the presence of amenities and characteristics such as stainless-steel appliances, elevator, laundry, concierge, and pool. Additionally, an aggregated feature sentiment_score was generated and included in the model, reflecting the emotional tone of the property descriptions.

Features Derived from the Proximity to Restaurants

To assess the impact of commercial infrastructure on rental prices, the coordinates of 344 Starbucks restaurants within Houston (Fig. 1) were used [32]. For the construction of new features, the authors matched the coordinates of real estate properties with Starbucks locations, applying the KD-tree algorithm for efficient nearest neighbour search and distance calculation in multidimensional space. This approach ensures high speed and accuracy when processing a large number of property–restaurant pairs.

Based on these calculations, the following features were generated:

•the number of Starbucks restaurants within radii of 1, 2, 3, 4, 6, 8, and 10 km from each property – these indicators reflect the density of outlets in the vicinity;

• the distance to the nearest Starbucks restaurant – as an indicator of minimum distance to a point of attraction, i.e., walkable accessibility;

• the average distance to the 2, 3, 4, 5, 6, 7, and 8 nearest restaurants – to evaluate the accessibility of outlets in the surrounding area.

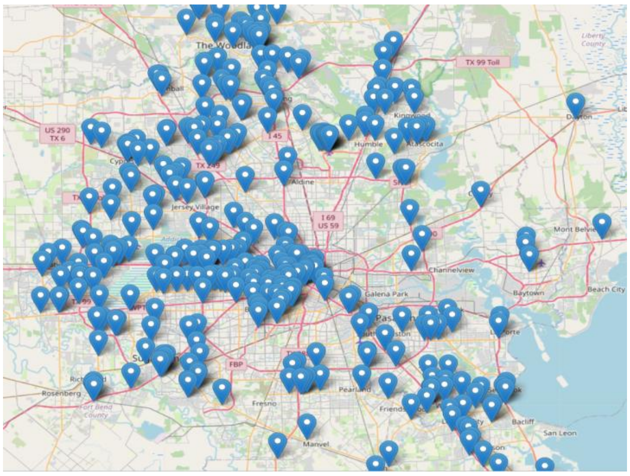

Fig. 1. **Map of Starbucks locations in Houston**
***Source*: prepared by the authors [32]**

Thus, the model integrated features that capture both the density and spatial distribution of Starbucks around each property, enabling the inclusion of infrastructural attractiveness and commercial environment in predicting rental prices.

Features Derived from Public Transport Stops

To construct features related to transport accessibility, geographic data on bus stops were obtained from the open-source platform OpenStreetMap.

Based on this data, the following characteristics were added to the dataset:

• the number of stops within radii of 100 m, 300 m, 500 m, 1 km, 2 km, 4 km, 6 km, 8 km, and 10 km from each property;

• the distance to the nearest bus stop;

• the average distance to the 2, 4, 6, and 8 nearest stops.

The selected distances reflect different levels of transport accessibility: 100-300 m correspond to walking distance to a stop for most residents, while larger radii (from 500 m to 10 km) make it possible to account for the overall provision of public transport in a given area. The use of these intervals enables the coverage of multiple scales of analysis and helps avoid feature redundancy caused by excessively small step sizes.

The calculation of distances between properties and stops was performed using their coordinates and efficient spatial structures such as the KD-tree

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

algorithm, which allowed for rapid nearest-neighbour searches in large datasets. This accelerated processing and ensured scalability when expanding the dataset. The resulting features describe the transport infrastructure and allow for an analysis of its influence on rental property prices.

Features Derived from Urban Night-Time Illumination

To extract features reflecting the level of night-time activity and infrastructural saturation of neighbourhoods, satellite images of the city of Houston at night were used, provided by NASA Earth Observatory [33]. High-resolution images were pre-aligned with geographic coordinates using QGIS tools.

The following processing steps were then carried out sequentially [34]:

• the image was converted to greyscale, after which contrast and brightness were adjusted to highlight brightly illuminated areas;

• binarization and contour detection of bright regions were applied, with their centres determined as the pixel coordinates of the brightest points;

• these pixel coordinates were transformed into geographic coordinates, taking into account the boundaries of the image and map dimensions;

• the resulting points were further clustered using the DBSCAN algorithm, in order to merge nearby light centres into groups and determine their centroids as global light clusters.

Based on the geographic coordinates of the final light centres, new features were generated:

• the number of clusters within radii of 0.1, 0.3, 0.5, 1, 2, 4, 6, 8, and 10 km from each property;

• the distance to the nearest light centre;

• the average distance to the 2, 4, 6, and 8 nearest light clusters.

These features reflect the degree of urban activity, potentially associated with the commercial or social intensity of the neighbourhood.

Feature Selection and Model Training Methodology

The LightGBM gradient boosting algorithm was employed as the primary predictive model. This model is built on an ensemble of decision trees trained sequentially, with each new element correcting the errors of the previous ones. Unlike traditional boosting, LightGBM applies a leaf-wise tree growth strategy, which improves accuracy on complex datasets. The model supports missing value handling, is efficient with a large number of features, and is robust to multicollinearity – a critical property when working with heterogeneous data from diverse sources.

The authors applied recursive feature elimination with cross-validation (RFECV) to enhance the generalization performance of the model and remove redundant features at the beginning phase. The method of elimination removes the least significant features in a repetitive manner, and it re-trains the model at every step, with cross-validation measures to evaluate the performance of the model. This would allow the researcher to make a good choice of features that would generate a compromise between the complexity of the model and precision.

For automatic hyperparameter tuning, the Optuna framework, a modern tool based on Bayesian optimization, was used. Unlike GridSearch or RandomSearch, Optuna enables intelligent exploration of the parameter space and uses strategies to stop ineffective attempts early, thereby saving resources and increasing the likelihood of finding the best configuration.

The main hyperparameters of the LightGBM model included the number of estimators, learning rate, maximum tree depth, and feature fraction. The optimisation process was performed using Optuna with early stopping based on validation loss, allowing efficient exploration of the parameter space.

The main stages of model development included:

• sequential integration of the base dataset with additional groups of heterogeneous features derived from illumination, textual descriptions, and data on restaurants and public transport stops, in order to gradually assess the contribution of each group and preserve interpretability of the results.

• use of LightGBM as the main gradient boosting model, capable of efficiently handling heterogeneous and incomplete tabular data;

• selection of optimal features using RFECV based on decision trees – to eliminate uninformative or interdependent variables and improve model robustness;

• hyperparameter optimisation with Optuna, which allowed for an improvement in model accuracy without a substantial increase in computational costs.

To justify the selection of the LightGBM model, a comparative analysis with alternative approaches was conducted using the baseline feature set. All models were evaluated using default hyperparameters without additional tuning to ensure a fair and unbiased comparison of their inherent predictive capabilities. It should be noted that this comparison was performed outside the full modelling pipeline and does not include feature selection or hyperparameter optimisation. Therefore, the obtained results reflect the relative performance of the algorithms under simplified conditions and are not directly comparable to the results reported in subsequent sections, where a complete pipeline with cross-validation and optimisation procedures is applied.

The results in Table 2 demonstrate that LightGBM provides the best predictive performance in terms of both RMSE and R^2, which justifies its selection as the primary model for further experiments.

Table 2. **Model comparison results**

Model	RMSE	R^2
Linear Regression	607.10	0.559
Random Forest	346.67	0.856
LightGBM	331.11	0.869

Source: **compiled by the authors**

This methodology ensured both high prediction quality – through effective feature selection combined with model hyperparameter optimisation – and transparency, thanks to the step-by-step introduction of thematic feature blocks and separate evaluation of their contribution.

Algorithm for Result Interpretation

After training the model, the next stage was the interpretation of its decisions. For this purpose, modern methods combining visualisation and quantitative feature contribution assessment were used. Two complementary tools were applied, providing both local and global levels of analysis.

1. SHAP (SHapley Additive ExPlanations) using TreeExplainer to determine how much of an individual's prediction can be attributed to each feature. Shapley values from game theory are used to assign credit to players (features). It determines the contribution of each feature to the final prediction. By calculating the difference between the predicted value and the baseline (expected) value, you can see how much a feature contributed to increasing or decreasing the prediction. For example, if a property has a larger area than average, SHAP will show that this factor shifted the final price upwards. In addition to pointwise analysis (local interpretations), SHAP values can be aggregated to provide global insights into feature importance.

2. Partial Dependence Plots (PDP) were applied as a visualisation tool for the influence of features on the target variable. Unlike SHAP, which captures the feature contributions considering interactions and the context of a specific object, PDP makes it possible to average the influence of one feature across all observations, smoothing the impact of other variables. This allowed for a detailed characterisation of feature influence, such as threshold effects, saturation zones, or non-linear relationships.

Thus, the combined application of SHAP and PDP provided a comprehensive view of model behaviour: first, identifying which features are important and how they act in each individual case; then, clarifying the shape and structure of their influence at a global level. This strategy ensured high model transparency and the possibility of formulating well-grounded practical recommendations.

The interpretation stage described above represents the final component of the overall modelling pipeline. The complete workflow includes data collection, preprocessing, feature engineering, feature selection using RFECV, model training and hyperparameter optimisation with LightGBM and Optuna, followed by model evaluation and interpretation.

The full data processing and modelling pipeline is presented in Fig. 2.

Results and Discussion

To evaluate the predictive performance of the model, standard regression metrics were used, including Mean Squared Error (MSE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). MSE reflects the average squared deviation between predicted and actual values, RMSE represents the error in the same units as the target variable, and R^2 indicates the proportion of variance in the target variable explained by the model.

Unlike the initial model comparison presented earlier, which was performed using default hyperparameters, the following experiments were conducted within the full modelling pipeline, including feature selection (RFECV) and hyperparameter optimisation with Optuna under cross-validation, providing a more realistic estimate of model performance. The impact of different feature groups was assessed through a step-by-step integration of heterogeneous feature sets, with the model trained sequentially by adding textual, visual, and geospatial features to the baseline dataset.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

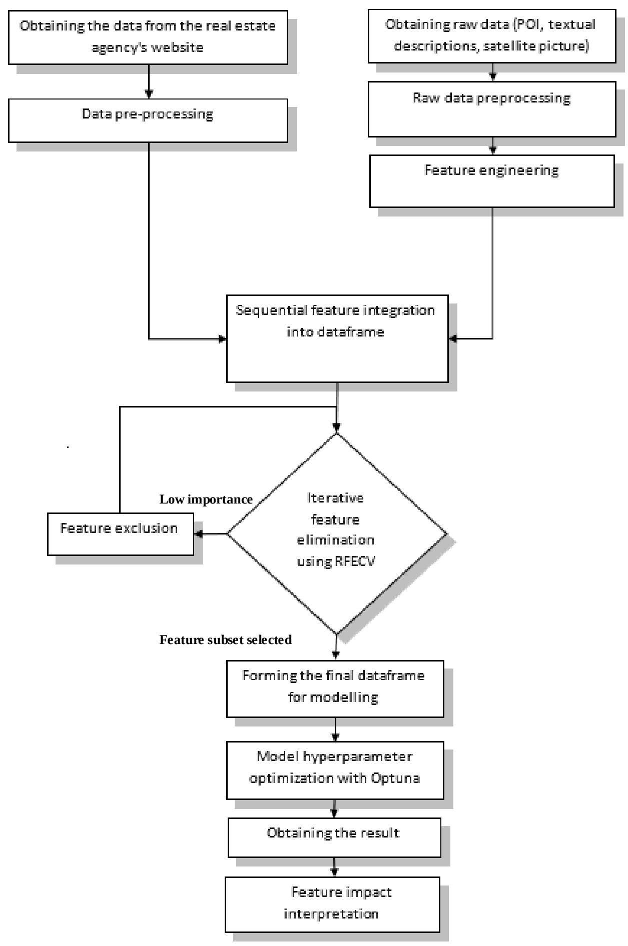

Fig. 2. **Scheme of the model for assessing real estate operations projects values**
***Source*: compiled by the authors**

The results in Table 3 indicate that the inclusion of textual features leads to the most significant improvement in prediction accuracy, reducing RMSE from 374 to 348 and increasing R^2 from 0.818 to 0.834. Additional feature groups, such as lighting and commercial infrastructure, also improve model performance, although to a lesser extent. Transport accessibility features show a weaker contribution, which may be related to local characteristics of the studied area.

Table 3. **Model performance depending on the feature set**

Feature set	RMSE	R^2
Baseline	374	0.818
Text features	348	0.834
Lighting features	351	0.839
Commercial infrastructure	354	0.837
Transport accessibility	365	0.826

***Source*: compiled by the authors**

Analysis of prediction errors showed that the largest deviations occur for atypical properties with rare combinations of features, such as unusually small or large areas, or uncommon textual descriptions. This indicates that the model performs less reliably in cases that are underrepresented in the training data.

This section presents the key results of the analysis aimed at interpreting the factors influencing rental property prices. The examination was carried out sequentially for each group of features: first, the textual descriptions of properties were studied (Fig. 3), followed by spatial proximity to food service establishments, then the density and accessibility of public transport, and finally the indicators of urban night-time illumination. This order reflects the transition from subjective characteristics to objective properties of the urban environment.

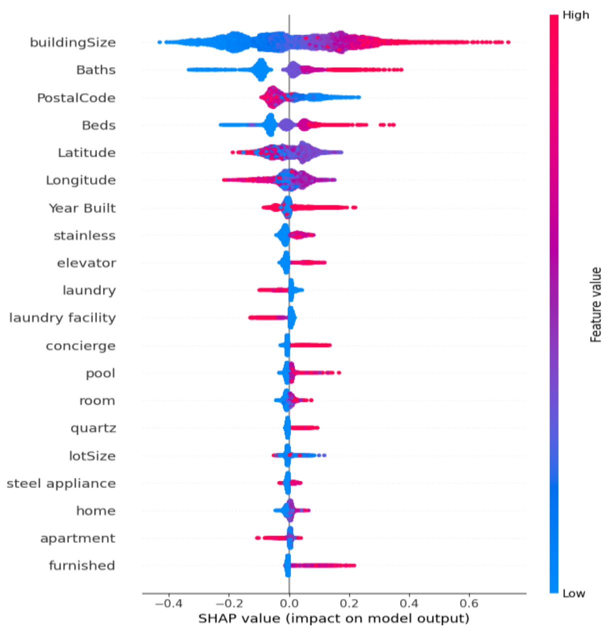

Fig. 3. **SHAP value impact**
***Source*: compiled by the authors**

In total, at the final stage of the study, the influence of 65 new and 9 baseline features on the target variable “Price” was assessed, and the authors decided to focus on the most influential features from each subgroup to avoid overloading the analytical section.

Fig. 3 shows the SHAP value diagram for the 20 most significant features used in the LightGBM model, including both baseline and text-derived ones. The X-axis represents SHAP values, reflecting the strength and direction of a feature’s impact on the prediction (positive values increase the forecast, negative values decrease it). The features are ordered by decreasing overall importance (vertically). The colour scale from blue (low feature value) to pink (high) makes it possible to analyse how variations in feature values are associated with changes in the model's output variable.

With reference to the table above, the following observations can be made:

• the building size, which is the most significant feature in the table; the larger the building size, the higher the rental price will be;

• the number of bedrooms and bathrooms is also positively correlated with the rental price. The more baths and beds, the higher the rental price;

• the latitude/longitude/postal code all exhibit a significant impact; however, the direction of their impact can be either positive or negative, depending on their values. The reasons lie in the location (latitude & longitude) as well as the postal code (example: two neighbouring homes situated at locations with slightly different latitudes, longitudes and/or postal codes);

•Year Built: more recent construction generally corresponds to a positive contribution, though the effect is moderate;

• lotSize: has limited influence, with a predominance of weakly negative effects as the value increases;

• concierge, pool, elevator, laundry facility, stainless, steel appliance: the presence of these terms in property descriptions is associated with an increase in the predicted value. These features are typically linked to higher standards of comfort and service;

• laundry, room, apartment, home: the presence of these terms may slightly reduce the final estimate or have little effect. This may be due to their frequent use in describing more standard or mass-market properties;

• furnished: the feature's influence varies depending on its value, without a clear consistent effect. Both positive and negative SHAP values are visible on the chart;

• quartz: demonstrates a bidirectional influence, suggesting differences in the perception of this feature depending on context (e.g., when mentioned in properties of different classes).

Thus, among the textual features, it is possible to distinguish a group of terms consistently associated with more expensive housing (e.g., pool, concierge, elevator), as well as words with uncertain or neutral contributions. The SHAP chart shows the importance of the features and which direction they are going in, which is very important when it comes to understanding models used in real estate appraisal tasks.

The graph in Fig. 4 demonstrates the actual link between a building's square footage and its corresponding (potential) rental value (in logarithmic terms) as computed by the PDP method. As the graph indicates, the rent increases dramatically until about 1300 sq. ft. as a function of the basic necessity of living space Beyond this, a saturation effect is observed: after 1,300-1,500 square feet, the growth becomes more moderate, and starting from 1,750 square feet, the influence of area on price stabilises almost completely. This form of dependency is typical for rental markets, where the importance of additional space decreases once a certain comfort threshold has been reached.

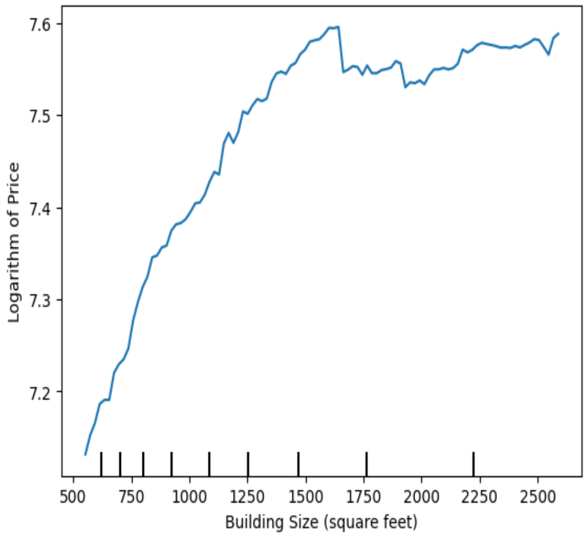

Fig. 4. **PDP Building size vs Price**
***Source*: compiled by the authors**

Fig. 5 shows Partial Dependence Plots (PDPs) that depict the impact of 4 features on rental prices. Based on these pictures:

• latitude and rental prices demonstrate a nonlinear relationship. Rental prices climb as the latitude of properties increases toward approximately 29.75, a point that likely represents more fashionable or central areas of the city, and then begin to fall as latitude continues to increase past this threshold;

• elevator: increase rent, with the influence of elevators growing stronger once they exceed a certain threshold (e.g., 0.03) in the model. This reflects the higher demand for properties with elevators, particularly in multi-storey buildings;

• pool: also has a positive impact, with the dependency remaining almost linear, the higher the importance of this feature in the description, the higher the predicted rental price;

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

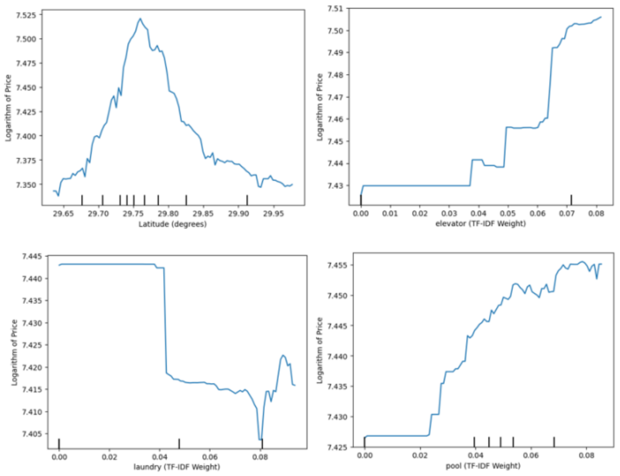

Fig. 5. **Features vs Price**
***Source*: compiled by the authors**

• laundry: by contrast, shows the opposite effect, starting from a certain term weight (around 0.03), a decrease in rental price is observed. This may be due to the fact that the mention of laundry is often found in more budget-oriented listings, or carries neutral or even negative informational value.

Fig. 4 and Fig. 5 complement the results obtained through SHAP analysis, allowing for a more precise interpretation of the shape of feature influence and for the identification of threshold effects and saturation zones. The combined use of SHAP and Partial Dependence provides both local and global explanations, thereby deepening the understanding of the model's behaviour.

In a similar way, using Partial Dependence Plots, the most influential features from the group related to the location of Starbucks restaurants were analysed (Fig. 6), which made it possible to visualise the impact of spatial density and accessibility of outlets on rental price predictions:

• the plot of the number of Starbucks restaurants within a 2 km radius shows an increase in the logarithm of price up to around 10 outlets, after which the effect weakens — likely due to saturation;

• the plot for outlets within a 10 km radius displays a more stable dependency: the more restaurants there are, the higher the price, with a particularly noticeable jump after 70 outlets, which may indicate proximity to business centres;

• The distance to the nearest restaurant exhibits a non-linear dependency: the effect is strongest at moderate distances, whereas being too close or too far does not have a significant positive influence;

• the average distance to the 7 nearest restaurants shows a clear downward trend – the closer the outlets on average, the higher the predicted rental price.

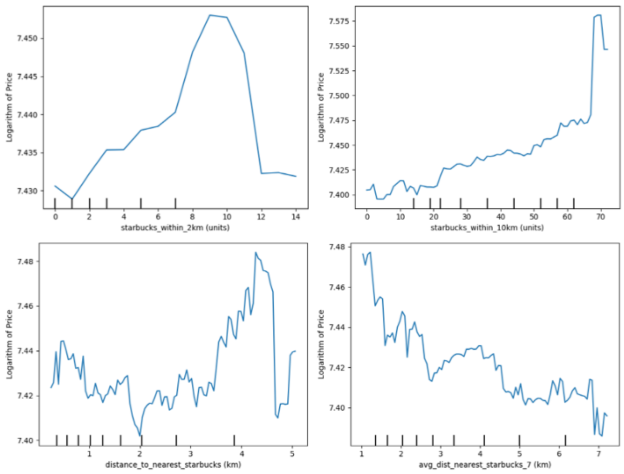

Fig. 6. **Restaurants-based features vs Price log**
***Source*: compiled by the authors**

Thus, the influence of restaurants on price is non-linear: moderate density and distance are associated with rising prices, but the effect saturates when outlets become too numerous, while overall closer average proximity consistently increases rental value.

Fig. 7 presents Partial Dependence Plots of the predicted logarithm of rental prices against the number of bus stops at various distances, identified as the most informative features in the next subgroup:

• bus_within_0.1km – a weak but steady increase is observed: as the number of stops within a 100 m radius grows, the price rises, which may indicate a preference for locations with minimal walking distance to public transport;

• bus_within_10km – a pronounced non-linear increase occurs after a certain threshold, which may reflect infrastructural saturation and building density: prices are higher in areas with a well-developed transport network;

• bus_within_8km – the dependency is also positive, but with saturation followed by a decline after the peak, which may indicate the effect of overcongestion.

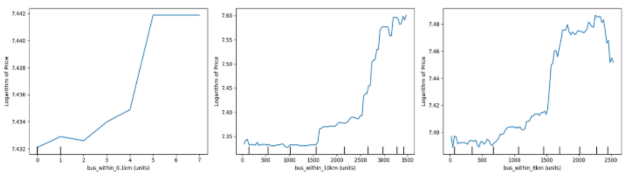

Fig. 7. **Transport based features vs Price log**
***Source*: compiled by the authors**

Thus, the overall trend shows a positive influence of the number of stops on rental prices, especially at moderate values, although threshold and saturation effects typical of complex urban environments are also observed.

Fig. 8 presents Partial Dependence Plots between the predicted property price (in logarithmic scale) and features reflecting the spatial proximity of urban illumination clusters, identified as the most influential in the final subgroup of features:

• avg_dist_nearest_clusters_8: shows that as the average distance to the eight nearest light clusters increases, the price tends to decrease, indicating a positive effect of proximity to illuminated areas;

• for the most part, distance to nearest cluster(s) speaks to a fairly volatile distance range, but there is still an overall benefit to being close to light, i.e., being near a city;

• clusters within 10km show that an area is undergoing an increase in value (moderate) as the number of clusters within 10km grows, until a certain point, and then the area's value reaches saturation.

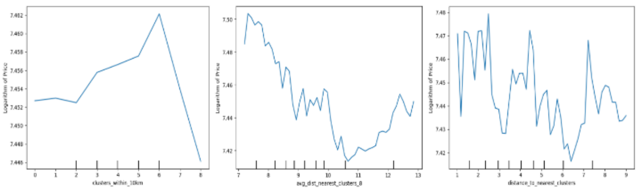

Fig. 8. **Urban lights-based features vs Price log**
***Source*: compiled by the authors**

These relationships are further substantiating evidence that night time light is an indicator of urban infrastructure and activity and is associated with the attractiveness of a location.

As an additional method for interpreting the model's prediction for a specific property with unique baseline feature values, a waterfall chart (Fig. 9) of their contributions to the overall prediction was used.

The following can be noted:

• the model's baseline value is approximately 1844.2, which reflects the average prediction across the entire training sample;

• the final prediction for this property is 1491.7, which is significantly below the average level;

• features that reduced the predicted price:

– building size (buildingSize = 780): −288.6. The small property size substantially decreases the predicted value;

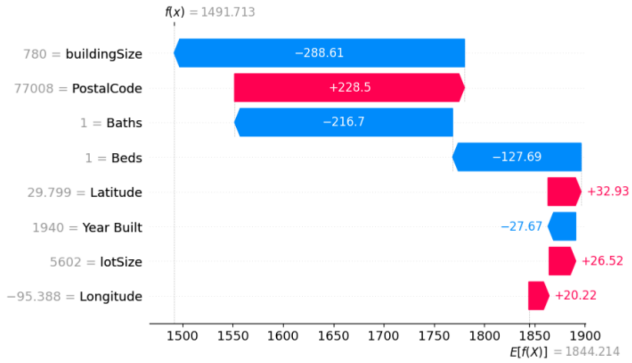

Fig. 9. **SHAP waterfall plot of basic features**
***Source*: compiled by the authors**

– number of bathrooms (Baths = 1): −216.7. Having only one bathroom negatively influences the estimate;

– number of bedrooms (Beds = 1): −127.7. Limited living space lowers the price;

– year built (Year Built = 1940): −27.7. The old construction date has a negative effect;

• features that increased the predicted price:

– postal code (PostalCode = 77008): +228.5. The advantageous location adds a significant positive adjustment;

– latitude (Latitude = 29.799): +32.9. The geographical position has a moderate positive effect.

– lot size (LotSize = 5602): +26.5. A larger lot size contributes positively to the value;

– longitude (Longitude = −95.388): +20.2. This also provides a small positive contribution.

The SHAP waterfall graph depicts the way that each individual feature affects the model output, either positively or negatively, in relation to the base value, to explain the contribution of the driving feature to the forecast of the model. This approach enables the model to be explainable at individual observation levels. This interpretation method works well when trying to forecast the future value of real estate. The two tools that we have used to find the most significant attributes are SHAP (SHapley Additive exPlanations) and Partial Dependence Plots, which show the way these attributes affect the predictions of the model in detail.

This approach not only ensured both quantitative and visual interpretation but also offered a glimpse inside the algorithm's "black box", making the model's behaviour more transparent. The methodology can be considered a complement to existing analytical tools, while giving the researcher a deeper understanding of the factors shaping the final price of a property.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

CONCLUSIONS

In this study, a comprehensive method was proposed to improve the accuracy of predictions and to analyse the features influencing rental property prices using interpretable machine learning models. The approach combines the calculation of SHAP values with the visualisation of Partial Dependence, which makes it possible not only to identify the most important features but also to interpret the nature of their influence on price, including the detection of non-linearities, threshold effects, and saturation zones.

The conducted research, based on a dataset of rental property listings in Houston (USA), covered a wide range of features: numerical, categorical, textual (processed through TF-IDF), as well as spatial and geo-analytical characteristics. It was established that such features as property size, number of bathrooms, latitude, the presence of an elevator, a pool, or terms related to interior (e.g., furnished) have a significant influence on price. Geospatial features related to the density of points of interest (restaurants, public transport stops) and proximity to light clusters allowed for a more precise consideration of the geographic characteristics of real estate objects.

Although the experiments were conducted using data from Houston, the proposed methodology is not limited to a specific city. The feature engineering approach and interpretation framework can be transferred to other real estate markets, while specific feature effects may vary depending on local socio-economic conditions.

The analysis of results makes it possible to categorise features by the degree of their potential generalisability:

• basic property characteristics (size, number of rooms, year built, etc.) can be regarded as relatively universal factors influencing price across most markets, although the magnitude of the effect may vary;

• infrastructural features (presence of Starbucks restaurants, accessibility of public transport) largely reflect local socio-economic conditions and require re-analysis when transferred to other cities;

• geospatial indicators, such as measures of illumination and proximity to clusters of activity, occupy an intermediate position: their qualitative influence is confirmed by the study, but the quantitative effect depends on the specifics of the urban environment in a given region.

The methodology has significantly improved the interpretability of the model and provided a deeper understanding of the factors that determine prices in the rental market. Unlike traditional approaches, the proposed solution allows not only to quantitatively assess the impact of characteristics, but also to visually track complex dependencies, which increases the practical value of the results. This approach can be effectively applied in automated valuation models (AVMs), in urban studies, and to support informed decision-making by investors and developers.

Additionally, the versatility of the proposed methodology makes it applicable in related domains – for example, in assessing the market value of automotive, maritime, and other types of transport [35], [36], where textual descriptions, technical parameters, and geolocation features can also be utilised.

REFERENCES

1. Lorenz, F., Willwersch, J., Cajias, M. & Fuerst, F. "Interpretable machine learning for real estate market analysis". *Real Estate Economics*. 2022. 51 (4), https://www.scopus.com/pages/publications/85132747566. DOI: https://doi.org/10.1111/1540-6229.12397.

2. Neves, F. T., Aparicio, M. & de Castro Neto, M. "The impacts of open data and explainable AI on real estate price predictions in smart cities". *Applied Sciences*. 2024; 14 (5): 2209. DOI: https://doi.org/10.3390/app14052209.

3. Hamolia, V., Melnyk, V., Zhezhnych, P. & Shilinh, A. "Intrusion detection in computer networks using latent space representation and machine learning". *International Journal of Computing*. 2020; 19 (3): 442–448. DOI: https://doi.org/10.47839/ijc.19.3.1893.

4. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. "Explainable AI: A review of machine learning interpretability methods". *Entropy*. 2021; 23 (1): 18. DOI: https://doi.org/10.3390/e23010018.

5. Zheng, Q., Seto, K. C., Zhou, Y., You, S. & Weng, Q. "Nighttime light remote sensing for urban applications: Progress, challenges, and prospects". *ISPRS Journal of Photogrammetry and Remote Sensing*. 2023; 202: 125–141. DOI: https://doi.org/10.1016/j.isprsjprs.2023.05.028.

6. Dong, B., Zhang, R., Li, S., Ye, Y. & Huang, C. "A meta-analysis for the nighttime light remote sensing data applied in urban research: Key topics, hotspot study areas and new trends". *Science of Remote Sensing*. 2025; 11: 100186, https://www.scopus.com/pages/publications/85212578883. DOI: https://doi.org/10.1016/j.srs.2024.100186.

7. Han, J., Woo, A. & Lee, S. "Effects of neighborhood streetscape on the single-family housing price: Focusing on nonlinear and interaction effects using interpretable machine learning". *PLOS ONE*. 2025. DOI: https://doi.org/10.1371/journal.pone.0323495.

8. Lipyanina, H., Maksymovych, V., Sachenko, A., Lendyuk, T., Fomenko, A. & Kit, I. "Assessing the investment risk of virtual IT company based on machine learning". In: Babichev, S., Peleshko, D., Vynokurova, O. (eds.). *Data Stream Mining & Processing*. *DSMP*. *Communications in Computer and Information Science*. 2020. 1158. *Springer*. Cham. DOI: https://doi.org/10.1007/978-3-030-61656-4_11.

9. Duhan, M. & Bhatia, P. K. "Software reusability estimation based on dynamic metrics using soft computing techniques". *International Journal of Computing*. 2022; 21 (2): 188–194. DOI: https://doi.org/10.47839/ijc.21.2.2587.

10. Lynnyk, R., Vysotska, V., Matseliukh, Y., Burov, Y., Demkiv, L., Zaverbnyj, A., Sachenko, A., Shylinska, I., Yevseyeva, I. & Bihun, O. "DDOS attacks analysis based on machine learning in challenges of global changes". *CEUR Workshop Proceedings*. MoMLeT+DS 2020 Modern Machine Learning Technologies and Data Science Workshop. 2020: 159–171. – Available from: https://ceur-ws.org/Vol-2631/paper12.pdf. – [Accessed: Nov, 2025].

11. R., S., Kanavalli, A., Gupta, A., Pattanaik, A. & Agarwal, S. "Real-time DDoS detection and mitigation in software defined networks using machine learning techniques". *International Journal of Computing*. 2022; 21 (3): 353–359. DOI: https://doi.org/10.47839/ijc.21.3.2691.

12. Soprun, O., Bublyk, M., Matseliukh, Y., Andrunyk, V., Chyrun, L., Dyyak, I., Yakovlev, A., Emmerich, M., Osolinsky, O. & Sachenko, A. "Forecasting temperatures of a synchronous motor with permanent magnets using machine learning". *CEUR Workshop Proceedings*. *MoMLeT+DS 2020 Modern Machine Learning Technologies and Data Science Workshop*. 2020. p. 95–120. – Available from: https://ceur-ws.org/Vol-2631/paper8.pdf. – [Accessed: Nov, 2025].

13. Sirola, M. & Hulsund, J. E. "Machine-learning methods in prognosis of ageing phenomena in nuclear power plant components". *International Journal of Computing*. 2021; 20 (1): 11–21. DOI: https://doi.org/10.47839/ijc.20.1.2086.

14. Lipianina-Honcharenko, K., Wolff, C., Sachenko, A., Desyatnyuk, O., Sachenko, S. & Kit, I. "Intelligent information system for product promotion in internet market". *Applied Sciences*. 2023; 13 (17): 9585. DOI: https://doi.org/10.3390/app13179585.

15. Bhat, S. S., Selvam, V. & Ansari, G. A. "Predicting life style of early diabetes mellitus using machine learning technique". *International Journal of Computing*. 2023; 22 (3): 345–351. DOI: https://doi.org/10.47839/ijc.22.3.3230.

16. Nosov, P., Melnyk, O., Malaksiano, M., Mamenko, P., Onyshko, D., Fomin, O., Píštěk, V. & Kučera, P. "Machine learning-based semantic analysis of scientific publications for knowledge extraction in safety-critical domains". *Machine Learning and Knowledge Extraction*. 2025; 7 (4): 150. DOI: https://doi.org/10.3390/make7040150.

17. Nosov, P., Melnyk, O., Malaksiano, M., Shumylo, O., Onishchenko, O., Yarovenko, V., Zinchenko, S. & Popovych, I. "A unified fractal processing framework for normalized AIS and ECDIS ship trajectories". *Digital*. 2026; 6 (1): 11. DOI: https://doi.org/10.3390/digital6010011.

18. Bocharova, M. & Malakhov, E. "Improving information theory of context-aware phrase embeddings in HR domain". *Eastern-European Journal of Enterprise Technologies.* 2024; 5 (2 (131)): 53–60. DOI: https://doi.org/10.15587/1729-4061.2024.313970.

19. Bocharova, M. Y. & Malakhov, E. V. "CapStyleBERT: Incorporating capitalization and style information into BERT for enhanced resumes parsing". *Proceedings of the 13th International Conference on Software and Computer Applications (ICSCA).* 2024. p. 249–254. DOI: https://doi.org/10.1145/3651781.3651820.

20. Kalinina, I., Gozhyj, A., Vysotska, V., Malakhov, E., Gozhyj, V. & Tregubova, I. "System methodology of data analysis and preprocessing for solving classification problems". *Proceedings of the 19th*

IEEE International Conference on Computer Science and Information Technologies (CSIT). 2024. p. 1–6. DOI: https://doi.org/10.1109/CSIT65290.2024.10982630.

21. Iban, M. C. "An explainable model for the mass appraisal of residences: The application of tree-based machine learning algorithms and interpretation of value determinants". *Habitat International*. 2022; 128: 102660, https://www.scopus.com/pages/publications/85137121263.
DOI: https://doi.org/10.1016/j.habitatint.2022.102660.

22. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. & Pedreschi, D. "A Survey of methods for explaining black box models". *ACM Computing Surveys*. 2019; 51 (5): 93. DOI: https://doi.org/10.1145/3236009.

23. Apley, D. W. & Zhu, J. "Visualizing the effects of predictor variables in black box supervised learning models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2020; 82 (4): 1059–1086. DOI: https://doi.org/10.1111/rssb.12377.

24. Lundberg, S., Erion, G., Chen, H., et al. "From local explanations to global understanding with explainable AI for trees". *Nature Machine Intelligence*. 2020; 2 (1): 56–67, https://www.scopus.com/pages/publications/85133604939. DOI: https://doi.org/10.1038/s42256-019-0138-9.

25. Deng, L. & Zhang, X. "Boosting the accuracy of property valuation with ensemble learning and explainable artificial intelligence: The case of Hong Kong". *Annals of Regional Science*. 2025; 74: 32. DOI: https://doi.org/10.1007/s00168-025-01365-7.

26. Bachmann, S. "Interpretable machine learning for the German residential rental market – shedding light into model mechanics". *Aestimum*. 2025; 86: 25–46, https://www.scopus.com/pages/publications/105017049938. DOI: https://doi.org/10.36253/aestim-16351.

27. Wan, H., Roy Chowdhury, P. K., Yoon, J., et al. "Explaining drivers of housing prices with nonlinear hedonic regressions". *Machine Learning with Applications*. 2025; 21: 100707. DOI: https://doi.org/10.1016/j.mlwa.2025.100707.

28. Cook, T. R., Gupton, G., Modig, Z. & Palmer, N. M. "Explaining machine learning by bootstrapping partial dependence functions andShapley values". *The Federal Reserve Bank of Kansas City Research Working Papers*. 2021. DOI: https://doi.org/10.18651/RWP2021-12.

29. "Redfin. Real estate and rental listings". – Available from: https://www.redfin.com. – [Accessed: Feb, 2025].

30. Qaiser, S. & Ali, R. "Text mining: use of TF-IDF to examine the relevance of words to documents". *International Journal of Computer Applications*. 2018; 181 (1): 25–29. DOI: https://doi.org/10.5120/ijca2018917395

31. Bushuyev, S. D., Bushuiev, D., Kravtsov, D., Poletaev, N. I. & Malaksiano, M. "Machine learning model for house price predicting based on natural language text data analysis". *CEUR Workshop Proceedings*. 2024; 3711: 319–332. – Available from: https://ceur-ws.org/Vol-3711/paper20.pdf – [Accessed: Mar, 2025].

32. "Starbucks. Starbucks store locator". – Available from: https://www.starbucks.com/store-locator. – [Accessed: Feb, 2025]

33. "NASA Earth Observatory. Nighttime light satellite data". – Available from: https://earthobservatory.nasa.gov. – [Accessed: Feb, 2025]

34. Bushuyev, S. D., Bushuiev, D., Poletaev, N. I., Malaksiano, M. & Kravtsov, D. "A machine learning method for real estate operation projects forecasting". *CEUR Workshop Proceedings. IT Project Management*. 2024. p. 77–84. DOI: https://doi.org/10.23939/IW_itpm2024.077.

35. Kobets, V. M., Shlyakhtenko, I. D., Zinchenko S. M. "Development of a product pricing algorithm using RFM strategy for user cohorts using machine learning methods". *Herald of Advanced Information Technology*. 2025; 8 (4): 497–509. DOI: https://doi.org/10.15276/hait.08.2025.32.

36. Bilak Y. Y. "Development of a Combined Model for Analyzing Gas Mixtures using Machine Learning Methods". *Applied Aspects of Information Technology*. 2025; 8 (1): 24–37. DOI: https://doi.org/10.15276/aait.08.2025.2.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 10.04.2026
Received after revision 12.06.2026
Accepted 19.06.2026

DOI: https://doi.org/10.15276/hait.09.2026.21
УДК 004.8

Підвищення точності та інтерпретованості прогнозів цін на нерухомість за допомогою методів машинного навчання

Кравцов Дмитро Євгенійович [1)]
ORCID: https://orcid.org/0009-0005-0305-544X; dmkravtsov@gmail.com. Scopus Author ID: 59197585500
Полєтаєв Микола Іванович [1)]
ORCID: https://orcid.org/0000-0002-1340-582X; Poletaev@ukr.net. Scopus Author ID: 6603897743
Горжеланчик Пйотр [2)]
ORCID: https://orcid.org/0000-0001-9662-400X; Piotr.gorzelanczyk@ans.pila.pl. Scopus Author ID: 22979733400
Соколовський Едгар [3)]
ORCID: https://orcid.org/0000-0002-0770-4225; edgar.sokolovskij@vilniustech.lt. Scopus Author ID: 55903092900
[1)] Одеський національний морський університет, вул. Мечникова, 34. Одеса, 65029, Україна
[2)] Державний університет прикладних наук імені Stanislaw Staszic у Пілі, вул. Podchorążych, 10. Піла, 64-920, Польща
[3)] Вільнюський технічний університет імені Гедимінаса. Saulėtekio al. 11. Вільнюс, 11, Saulėtekio al. Вільнюс, LT-10223, Литва

АНОТАЦІЯ

Актуальність: зумовлена зростаючою потребою у застосуванні методів машинного навчання в задачах оцінювання нерухомості, де важливо не лише отримувати точні прогнози, але й розуміти фактори, що впливають на формування ціни, особливо у процесах прийняття рішень у сфері урбаністики та девелопменту. **Мета:** метою статті є розроблення та оцінювання підходу, що одночасно підвищує точність прогнозування та забезпечує інтерпретованість моделей машинного навчання при роботі з гетерогенними джерелами даних. **Завдання:** завдання дослідження включають аналіз різних груп ознак, а саме текстових описів об'єктів, просторових характеристик, зокрема відстаней до об'єктів інфраструктури, а також візуальних ознак, отриманих на основі супутникових даних нічного освітлення, і оцінювання їхнього спільного впливу на якість моделі. **Методи:** у роботі використано метод градієнтного бустингу над деревами рішень на основі алгоритму Light Gradient Boosting Machine, побудову просторових ознак із застосуванням метрик відстані, методи обробки текстових даних, а також інструменти інтерпретації на основі значень Шеплі та аналізу часткової залежності для виявлення впливу ознак на результат прогнозування. **Наукова новизна:** полягає в інтеграції гетерогенних ознак різної природи в єдину модельну схему та поєднанні взаємодоповнюючих методів інтерпретації, що дозволяє виявляти нелінійні залежності та ефекти взаємодії між змінними. **Практична значимість:** запропонований підхід може бути використаний в автоматизованих системах оцінювання нерухомості, урбаністичних дослідженнях та системах підтримки прийняття рішень, забезпечуючи більш прозоре розуміння механізмів формування ціни. **Результати:** результати показали, що запропонований підхід на основі моделі Light Gradient Boosting Machine забезпечує високу якість прогнозування. Для базового набору ознак модель досягає середньоквадратичної помилки 374 та коефіцієнта детермінації 0.818. Інтеграція гетерогенних груп ознак додатково покращує якість моделі, зменшуючи середньоквадратичну помилку до 348 та підвищуючи коефіцієнт детермінації до 0.839. Отримані результати також підтверджують, що текстові та візуальні ознаки відіграють важливу роль у виявленні нелінійних залежностей і порогових ефектів, які складно ідентифікувати традиційними методами. **Висновки:** запропонований підхід забезпечує підвищення точності та інтерпретованості моделей машинного навчання, що дозволяє здійснювати більш надійний та прозорий аналіз вартості нерухомості та розширює можливості практичного застосування таких моделей.

Ключові слова: машинне навчання; глибоке навчання; математична модель; статистичні дослідження; ймовірність; кореляція; прогнозування; нейронні мережі; інтелектуальний аналіз даних; обробка природної мови; трансформери; текстові ембедінги; транспорт; пояснюваний штучний інтелект; інтерпретованість моделей; інтерпретоване машинне навчання; інженерія ознак; градієнтний бустинг; системи підтримки прийняття рішень

ABOUT THE AUTHORS

Dmitriy E. Kravtsov - PhD candidate in Computer Science, Faculty of Computer Science. Odesa National Maritime University. 34, Mechnikov Str. Odesa, 65029, Ukraine
ORCID: https://orcid.org/0009-0005-0305-544X; dmkravtsov@gmail.com. Scopus Author ID: 59197585500
Research field: Machine learning; Explainable artificial intelligence; Feature engineering; Geospatial data analysis; Predictive modelling; Automated valuation models

Кравцов Дмитро Євгенійович - аспірант кафедри Комп'ютерних наук. Одеський національний морський університет, вул. Мечникова, 34. Одеса, 65029, Україна

Nikolay I. Poletaev - Doctor of Physical and Mathematical Sciences, Professor, Department of Technical Cybernetics and Information Technology named after Prof. R. V. Merkt. Odesa National Maritime University. 34, Mechnikov Str. Odesa, 65029, Ukraine
ORCID: https://orcid.org/0000-0002-1340-582X; Poletaev@ukr.net. Scopus Author ID: 6603897743
Research field: Combustion in dispersed systems; Dusty plasma; Nanostructure synthesis; Materials science; Machine learning; Data mining

Полєтаєв Микола Іванович - доктор фізико-математичних наук, професор кафедри Технічної кібернетики та інформаційних технологій імені проф. Р. В. Меркта. Одеський національний морський університет, вул. Мечникова, 34. Одеса, 65029, Україна

Piotr Gorzelańczyk - Assoc. Prof., PhD Eng., Head of Transport Department. Stanislaw Staszic State University of Applied Sciences in Piła. 10, Podchorążych Str. Piła, 64-920, Poland
ORCID: https://orcid.org/0000-0001-9662-400X; Piotr.gorzelanczyk@ans.pila.pl. Scopus Author ID: 22979733400
Research field: Transport; Road accidents; Safety; Forecasting; Optimisation

Горжеланчик Пійотр - доцент, доктор інженерних наук, завідувач кафедри Транспорту. Державний університет прикладних наук імені Станіслава Сташица в Пілі. вул. Підхорунжич, 10 Піла, 64-920, Польща

Edgar Sokolovskij - Doctor of Technical Sciences, Professor. Vilnius Gediminas Technical University. 11, Saulėtekio al. Vilnius, LT-10223, Lithuania
ORCID: https://orcid.org/0000-0002-0770-4225; edgar.sokolovskij@vilniustech.lt. Scopus Author ID: 55903092900
Research field: Machine Learning; Data Science; Intelligent Systems; Decision Support Systems

Соколовський Едгар - доктор технічних наук, професор. Вільнюський технічний університет імені Гедимінаса. 11, ал. Саулетекио. Вільнюс, LT-10223, Литва